

LOGISTIC REGRESSION GRADED ASSIGNMENT

```
/* data prepration*/
```

```
data NEW_CREDIT;
set credit;
if age<25 then age25=1; else age25=0;
if 26 le age le 40 then age40=1;else age40=0;
if 41 le age le 60 then age60=1;else age60=0;
if age gt 60 then age80=1;else age80=0;

if Monthly_Income<100 then Monthly_Income1=1;else Monthly_Income1=0;
if 101 le Monthly_Income le 1000 then Monthly_Income2=1;else
Monthly_Income2=0;
if 1001 le Monthly_Income le 3000 then Monthly_Income3=1;else
Monthly_Income3=0;
if 3000le Monthly_Income le 5000 then Monthly_Income4=1;else
Monthly_Income4=0;
if 5001 le Monthly_Income le 7000 then Monthly_Income5=1;else
Monthly_Income5=0;
if Monthly_income gt 7000 then Monthly_Income6=1;else Monthly_Income6=0;

if Gender="male" then Gender1=1;else Gender1=0;
if Gender="femae" then Gender2=1;else Gender2=0;

if Education="Graduate" then Education1=1;else Education1=0;
if Education="Matric" then Education2=1;else Education2=0;
if Education="PhD" then Education3=1;else Education3=0;
if Education="Graduate" then Education4=1;else Education4=0;
if Education="Post-Grad" then Education5=1;else Education5=0;
if Education="Professional" then Education6=1;else Education6=0;

if NumberOfDependents<1 then NumberOfDependents1=1;else
NumberOfDependents1=0;
if NumberOfDependents<2 then NumberOfDependents2=1;else
NumberOfDependents2=0;
if NumberOfDependents<3 then NumberOfDependents3=1;else
NumberOfDependents3=0;
if NumberOfDependents<4 then NumberOfDependents4=1;else
NumberOfDependents4=0;
if NumberOfDependents<5 then NumberOfDependents5=5;else
NumberOfDependents5=0;
if NumberOfDependents gt 5 then NumberOfDependents6=1;else
NumberOfDependents6=0;

if Occupation="Non-offi" then Occupation1=1;else Occupation1=0;
if Occupation="Officer1" then Occupation2=1;else Occupation2=0;
if Occupation="Officer2" then Occupation3=1;else Occupation3=0;
if Occupation="Officer3" then Occupation4=1;else Occupation4=0;
if Occupation="Self-Emp" then Occupation5=1;else Occupation5=0;

if Rented_OwnHouse="Ownhouse" then Rented_OwnHousel=1;else
Rented_OwnHousel=0;
```

LOGISTIC REGRESSION GRADED ASSIGNMENT

```
if Rented_OwnHouse="Rented" then Rented_OwnHouse2=1;else Rented_OwnHouse2=0;

if Region="Centr" then Region1=1;else Region1=0;
if Region="North" then Region2=1;else Region2=0;
if Region="East" then Region3=1;else Region3=0;
if Region="West" then Region4=1;else Region4=0;
if Region="South" then Region5=1;else Region5=0;
run;

/* Dividing data into training and validation dataset*/

proc surveyselect data=new_credit
method =SRS out=SAMP1 samprate=0.5 outall;
run;

data train validate;
set samp1;
if selected=0 then output train;
else if selected=1 then output validate;
run;

/*logistic model using backward option to eliminate insignificant variables*/

proc logistic data=new_credit descending;
model NPA_Status(event='1')= DebtRatio Education1 Education2 Education3
Education4 Education5 NumberOfOpenCreditLinesAndLoans NumberOfDependents1
NumberOfDependents2
NumberOfDependents3 NumberOfDependents4 NumberOfDependents5 Occupation1
Occupation2 Occupation3 Occupation4 Rented_OwnHouse1 Region1 Region2 Region3
Region4
NumberOfTime30_59DaysPastDueNotW NumberOfTime60_89DaysPastDueNotW
NumberOfTimes90DaysLate
NumberRealEstateLoansOrLines age25 age40 age60 / selection= backward;
ods output parameterestimates=model_2;
run;

/* final logistic model after using backward option*/

proc logistic data=train descending outmodel=dmm;
model NPA_Status(event='1')= Education2 Education3 Education5
NumberOfOpenCreditLinesAndLoans NumberOfDependents1
Occupation1 Occupation3 Rented_OwnHouse1 Region1 Region2 Region3 Region4
NumberOfTime30_59DaysPastDueNotW NumberOfTime60_89DaysPastDueNotW
NumberOfTimes90DaysLate
NumberRealEstateLoansOrLines age age25 age40 age60;
score out=dmp;
run;
```

LOGISTIC REGRESSION GRADED ASSIGNMENT

```
/* storing logistic equation into dmm dataset*/
```

```
proc logistic data=train descending outmodel=dmm;  
model NPA_Status(event='1')= Education2 Education3 Education5  
NumberOfOpenCreditLinesAndLoans NumberOfDependents1  
Occupation1 Occupation3 Rented_OwnHouse1 Region1 Region2 Region3 Region4  
NumberOfTime30_59DaysPastDueNotW NumberOfTime60_89DaysPastDueNotW  
NumberOfTimes90DaysLate  
NumberRealEstateLoansOrLines age age25 age40 age60;  
score out=dmp;  
run;
```

```
proc logistic data=validate descending outmodel=dmm;  
model NPA_Status(event='1')= Education2 Education3 Education5  
NumberOfOpenCreditLinesAndLoans NumberOfDependents1  
Occupation1 Occupation3 Rented_OwnHouse1 Region1 Region2 Region3 Region4  
NumberOfTime30_59DaysPastDueNotW NumberOfTime60_89DaysPastDueNotW  
NumberOfTimes90DaysLate  
NumberRealEstateLoansOrLines age age25 age40 age60;  
score out=dmp;  
run;
```

```
/*proc rank code*/  
proc rank data=dmp out= decile1  
groups=10 ties= mean;  
var p_1;  
ranks decile;  
run;
```

```
proc sort data=decile1;  
by descending p_1;  
run;
```

```
/* scoring the data/*
```

```
proc logistic inmodel=dmm;  
score= New_dataset_name out=score;  
run;
```

Summary of the model:-

After running the logistic model with backward option eliminating insignificant variables we observe:-

- 1.All the variables are significant in the range $P < 0.05$.
2. **Model Convergence Status** is satisfied.
- 3.The coefficient values are similar for both training as well as validation dataset.
- 4.**percent concordant value is 81.7** which is good.
- 5.BETA=0 TEST is also satisfied.
6. The lift curve shows us that our model is better than general prediction model.
- 7.The model equation can be further used for any other dataset.